

## SEARCH ENGINE OPTIMIZATION PAGE RANK ALGORITHM

DHRUVA S. PATEL<sup>1</sup> & KETAN SARVAKAR<sup>2</sup>

<sup>1</sup>Research Scholar, Department of Information Technology, Ganpat University, U V Patel College of Engineering,  
Kherva, Mehsana, Gujarat, India

<sup>2</sup>Ganpat University, U V Patel College of Engineering, Kherva, Mehsana, Gujarat, India

### ABSTRACT

*The world of technology and online presence of the giant size corporations have changed the definition of the business as well the marketing of the same. This is supposed to be so mainly because of the availability of the information of every business stack holder on the tips of fingers. However, but the instinct as it remains the same be it a layman or a multi-millionaire, he/she always prefers to buy something that is easily available requiring fewer efforts to search hither and thither. SEO is a technique helping the end-users to find what they want and that too from the choice of their place and brand. This amazing phenomenon takes place due to a wellset algorithm that determines the ranking of a particular website for a particular most used keyword on one hand and on the other hand the contents, graphics, website structure and the genuineness of the contents displayed on a web page. This research paper focuses on the core area of SEO- Keywords, Website Contents- On Page Optimization & Off-Page Optimization, and the Algorithm used by search engines to give a rank to a website. As an interesting experiment, I will also use the page rank method for UVPCE website data.*

**KEYWORDS:** WWW, Search Engine Optimization (SEO), Website Ranking, Ranking Algorithm, Inbound Link, Outbound Link, Web Pages & Page Rank

Original Article

**Received:** Mar 13, 2018; **Accepted:** Apr 07, 2018; **Published:** Jul 17, 2018; **Paper Id.:** IJCSETIRAUG20183

### INTRODUCTION

With the evolution of the Concept of Marketing Online in the 90's and the increasing trend of the WWW (World Wide Web) analysis compelled the business giants to think about the mechanism of the SEO (Search Engine Optimization). The SEO has become more challenging and imperative as well due to the countless web-pages, hyperlinks providing an enormous amount of the data relating to the interactions of the end-users. This information also throws light on the human behavior in terms of their tastes, preferences, choices, purchasing power which ultimately help the businesses to grow and expand to a larger extent. This is the reason why Search Engines plays a vital role in today's virtual businesses and hence the web page ranking is crucial to strengthen the virtual presence of any organization.

This research paper presents an in-depth analysis of the Page Rank methodology used by Google to rank different websites. This paper throws light on the significance of the Markov model used to create the link between different web pages and to construct a transition matrix for a graph and web pages.

SEO analysis considers the algorithm based on the contents evaluation and deep analysis of the hyperlink structure. Usually, a page rank can be found on the basis of prime three factors:

- Contents of the Website
- Weight
- The base score of the website pages

Or we can also say that in Weighted Page Algorithm, the focus is on in-link and out-link which primarily decides the relevance of a website page. Whereas HITS algorithm in IBM Clever categorizes the website pages into hubs to evaluate the weight of the website page.

The PR (Page Rank) is then implemented in Java and tested on two simple network graphs i.e. Google Web Graph & emails fetched by a research institution. This exercise helps in finding out the exactness and correctness of the implementation. For example, we can apply the page rank algorithm on UVPCE web pages and find out which web pages have desired page rankings. Now let us discuss how we can higher page rank for a larger database in less time.

## OPTIMIZING FACTORS

SEO analysis has lot more to do with the Optimizing factors. Optimizing factors are nothing but the diverse data which can be structured, Semi-structured, and even structured existing on a website page. Based on it, the analysis can also be broadly classified into the basic three Optimizing Areas:

- Keyword Analysis
- On page Optimization
- Off page Optimization

Let us discuss in detail the above three elements playing a vital role in the virtual strength of any website.

### Keyword Analysis

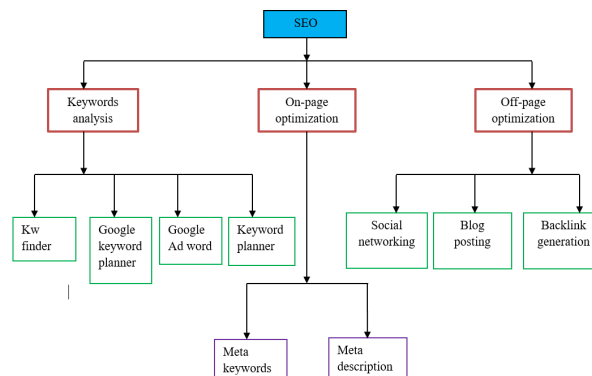
Key-word analysis is an important part of SEO as it would give you a set of the keywords used by the end-users from the targeted domain or territory you are focusing on. You can find the best-fitted set of keywords with the help of the tools are to ([to:https://adwords.google.com/intl/en\\_in/home/tools/keyword-planner/.](https://adwords.google.com/intl/en_in/home/tools/keyword-planner/))

With it, you can plan and place the relevant keywords on a website page that will boost the virtual presence of your website.

Keyword analysis is processes in which we find the keyword for our website using the google adword. Using this google adword we find the competition and search volume for our keyword. We can get an idea about keyword easily through the search result.

There are three types of keywords:

- **Long Tell:** In this type of keyword, we see the more search volume and short competition over the web.
- **Short Tell:** In this type of keyword find the more competition compare to all other keywords. Highly used in the search box.
- **Related:** In this type of keyword find the low competition. It is related to website product or content.



**Figure 1: SEO Factors**

### On-Page Optimization

This methodology includes placing of the keywords in title tag, keyword density, keyword in the meta tag, keywords in alt tag etc. This is a very first step followed by every prudent webmaster in order to make the web page more efficient in terms of visibility and rank. By applying this methodology, page rank definitely improves in the search engine and simultaneously gives a better satisfaction to the visitor as the result is what the end-user was intending.

On-page optimization can be followed using different ways to make a web-page more effective. For example, it can be done by

- Changing or modifying the Title
- Changing or modifying body text
- Changing or modifying the URL
- Changing or modifying the density of the keywords in a web-page. However, one needs to take care of the latest updates by Google in respect of the Key-words density.

If the on-page optimization is done with intuition and care, it is sure to bring about incredible results and productive traffic on your website page. We can say that on-page optimization is capable enough to be more specific in terms of the visitors visiting your website.

### Off-Page Optimization

Off-page optimization is nowadays becoming more and more challenging as it needs to align with the current Google updates and the smart use of the back-links. Off-page optimization focuses more on redirecting a visitor from other web-pages with the help of a back-link. The websites backed by more back-links are likely to get fruitful and resultant visitors. However, care has to be taken while Link-building.

We find diverse opinions stating that Back-links are almost outdated today but in fact, it is not the back-links but it is the Back-link plan fails. This is so mainly because of the placing a link without giving a thought to the whole plan and redirection cluster.

What makes the difference is the relevance of the back-link with the redirection page. For example, It would be a prudent practice to place a back-link in an article or a blog describing the significance of renewable sources of energy to a

webpage catering the services of IOT (Internet of Things). But the same will be good for nothing if the back-link is re-directed to a mere website development services.

Off-page optimization can be done by:

- Back link submission
- Social media management
- Blog submission
- Local listing
- Website sharing
- Social media bookmarking site
- Directory submission

## RANKING ALGORITHMS

Search engines algorithm is used to supply users associate results to get the information and retrieve the proper data as per user search box and present only relevant information in an ordered search. Many page rank algorithms rules are used to rank the online web pages and every algorithm has different mechanisms and parameters to calculate the relevant and important information of any data.

### Page Rank Algorithm

Page Rank is a calculation in which a numerical weight is appointed to a page as per its relative significance. It quantifies the significance of site pages. It utilizes approaching connection data to allow worldwide significance score to all pages on the web. A number of approaching connections from quality locales measures the prevalence of a page. It depends on the amount and nature of both inbound and outbound connections. Pages which have higher rank are most critical and it has opportunities to be recorded on web search tool's best outcome list. Page rank esteem is separated into levels 1-10 of which 10 speak to higher PR esteem implies that the page is more famous while page rank esteem 1 implies page isn't prominent.

The Page Rank Algorithm can be given in the following equation.

$$PR(A) = (1 - d)d + (PR \frac{T1}{C} (T1) + \dots PR \frac{Tn}{C} (Tn)) \quad (1)$$

Where d is damping coefficient and its value is 0.85.

PR(A)= page rank of webpage A,

C(tn)= number of outgoing links page tn.

### Weighted Page Rank Algorithm

Wenpu Xing and Ali Ghorbani introduced Weighted Page Rank calculation weighted page rank which is the alteration of the first PageRank calculation. Weighted page rank chooses the unmistakable quality of the pages by considering the significance of both in links and out links of the pages while doling out the rank score to the website pages.

This calculation does not equitably partition the page's rank among its out links rather than unique page positioning calculation and consequently gives high rank an incentive to the more mainstream pages. The prevalence is characterized by doling out weight esteems to the approaching and active connections are calculated by the following equations.

$$W_{(v,u)}^{in} = \frac{Iu}{\sum_{p \in R(v)} Ip} \quad (2)$$

Where  $Iu$  and  $Ip$ =number of in links of page and page  $p$ ,

$R(v)$  = reference page list of page  $v$ .

$$W_{(v,u)}^{out} = \frac{Ou}{\sum_{p \in R(v)} Op} \quad (3)$$

Where  $Ou$  and  $Op$  is the number of out links of page  $u$  and page  $p$ . By taking the popularity of the webpages into consideration the Weighted Page Ranking formula is given by the following equation:

$$PR(u) = (1 - d) + d \sum_{v \in B(u)} PR(v) W_{(v,u)}^{in} W_{(v,u)}^{out} \quad (4)$$

## PROPOSED ALGORITHMS

Another ranking algorithm named as PR algorithm. which utilizes Web Structure, web data and search engine optimization techniques to order to find the page rank of web pages over the large dataset. In which count the base scores of all the web pages or data and allout links and in links of web pages also compute the link weights and word count, also convert the unnormalized weight to the normalized weight. It uses the all outgoing link information from web pages. and give the important.

As an interesting side project, we decided to use the Page Rank algorithm that we have developed on webpages of Ganpat university website and Email from a large European research institution. We also made a significant adjustment in my text file: we reverse all the edges i.e. we convert all incoming edges in Figure shown to outgoing edges and vice versa. This is because it is more helpful to de ne a dependency relationship. We investigate which webpage is most important among the all web pages in the website using the page rank algorithm.

### Workflow of Algorithm:

- Get the web pages of websites from the input data.
- Assign all web pages rank at equals to one
- Initialize epsilon, base scores, score
- Compute the weights of in link and out links.
- count the number of links and word count, heading tag, alt tag
- Apply the proposed PR algorithm.
- Repeat the steps iteratively until important information of web pages is achieved.

In this proposed algorithm we are using the concept of simple page rank algorithm and take the result over different web pages over the web are as shown in the fig. It helps users to get the idea about only important and useful web pages in the website. So the user can easily select the useful web pages only.

## RESULTS ANALYSIS

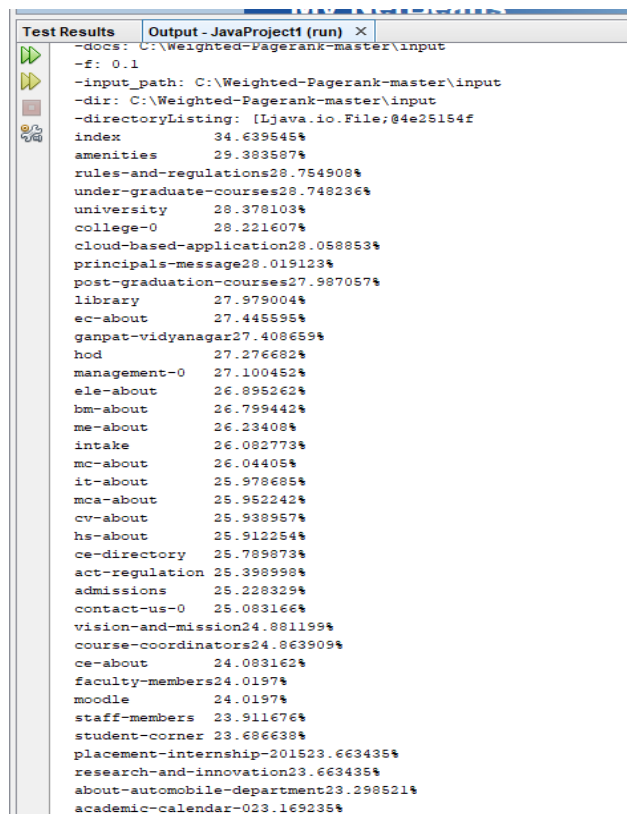


Figure 2: Important Web Pages in Decreasing Order

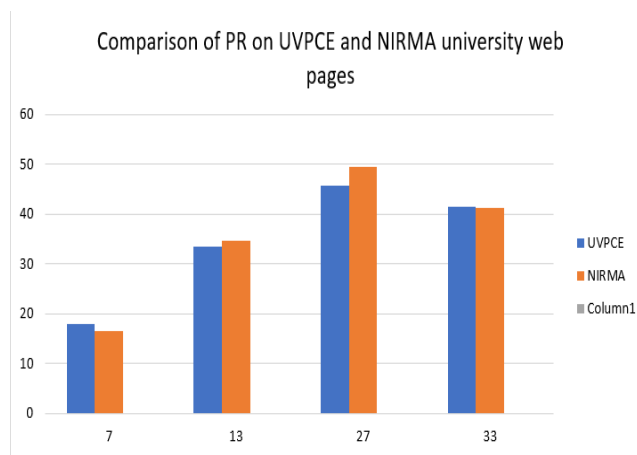


Figure 3: Comparison of Web Pages

The figure shows comparison and analysis chart of web pages of UVPCE and NIRMA university web pages with proposed page rank algorithm. It shows important web pages on the website.

Table 1

Web-Google	38.779305%
Email-EuAll	33.293865%
Com-youtube. top5000. cmt	27.92683%

Table 2

Name of Dataset	Google Programming Content
Nodes	875713
Edges	5105039
Number of triangles	13391903
Number of directed edges	4078

In this table, we are shown the information about google programming content. And compare with the you-tube and email content network dataset and find the more important webpages based on the nodes edges, In-links, and out-links of the webpage. So the user can easily get the more important web page Areas shown in table web-google has higher edges and nodes so it comes on first important webpage position Areas shown in the figure.

## CONCLUSIONS

In this paper page rank algorithm, calculation is proposed that take customer advancement ahead with the web page or website content for positioning the website or web pages using big data. In SEO there's forever chance for the development and its enhancement. SEO additionally has a few limitations and some disadvantages that don't allow the framework to be 100% right result. Stability of page rank can be improved. Further research should be focused on estimating the importance of each SEO technique and finding the proper information over the web.

## ACKNOWLEDGMENT

I thank my grateful faculty of my college at Ganpat university for this research, without which the present research could not be been completed I am especially indebted to my guide faculty Prof. Ketan Sarvakar, who have been Supportive and worked actively to provide me knowledge. I thankful to all my friends for their support during the research. I would like to thank my parents, whose love and guidance with me in whatever I pursue.

## REFERENCES

1. Patil Swati P. study of search engine Optimization. *International journal of advanced Research in computer science and software engineering, Haryana, India. volume 2, feb 2014 ISSN*
2. Vinit Kumar GunjanI. study of search engine Optimization with Google, On page Optimization, Off page optimization, *International journal of advanced Research in computer science L. B Nagar, Hyderabad, India, Haryana, India. volume 9, july 2014 ISSN:1694-0814.*
3. Wenpu Xing and Ali Ghorbani. *HITS, PageRank, Weighted PageRank, web mining University of New Brunswick Fredericton, IEEE 2017.*
4. Tuhena Sen. *Study simple page rank, HITS, page rank algorithm volume, amity university, Noida, India, 2017 IEEE*
5. Vinu jalswal. *A Review Paper on SEO based Ranking of Web Documents, Dept. of Computer Science & Engineering, GZS PTU Campus volume 2, july 2014 ISSN: 2277 128X*
6. Gauri Agarwal, Aparna Mane, Dept of computer engineering, search engine, *international journal of computer applications(0975) coulme 165-NO 1, May 2017.*
7. Kishor, K. "Comparative Studies On The Performance Parameters, Exhaust Emissions And Combustion Characteristics Of Modified Si Engine Over Conventional Engine."

8. Khanna, Nikhil, Nishant Behar, and Nishi Yadav. "Improved Algorithm for Dynamic Memory Allocation in Cloud Computing." (2016).
9. Nandnee jain, upendra Dwivedi, indore, Ranking web pages based on user interaction, ICACEA IMS engineering college, India , 978-4673@IEEE 2015
10. Sandhya verma, Amit kumar majhvar, upadated page rank of dynamically generated research author's pages, May 20-21 IEEE 2016
11. Chhanga, Devji, And Xitij Shukla. "Fossick: An implementation of federated search engine." *International Journal Of Computer Science Engineering And Information Technology Research (IJCSEITR)* 6.1 (2016): 69-78.
12. Kumar, Manoj. "Image Compression With Efficient Codebook Initilization Using Lbg-Optimization Algorithm."
13. Dr. Daya gupta, department of computer science and engineering, delhi Technolocal university, User Preference Based page ranking algorithm, ICCCA 2016
14. <https://en.wikipedia.org/wiki/PageRank>
15. <https://www.pagerank.net/pagerank-checker/>
16. A Novel User Preference and Feedback based Page Raking Technique. *IEEE* -2015.
17. [https://adwords.google.com/intl/en\\_in/home/tools/keyword-planner/](https://adwords.google.com/intl/en_in/home/tools/keyword-planner/)
18. [https://www.google.com/webmasters/#?modal\\_active=none](https://www.google.com/webmasters/#?modal_active=none)